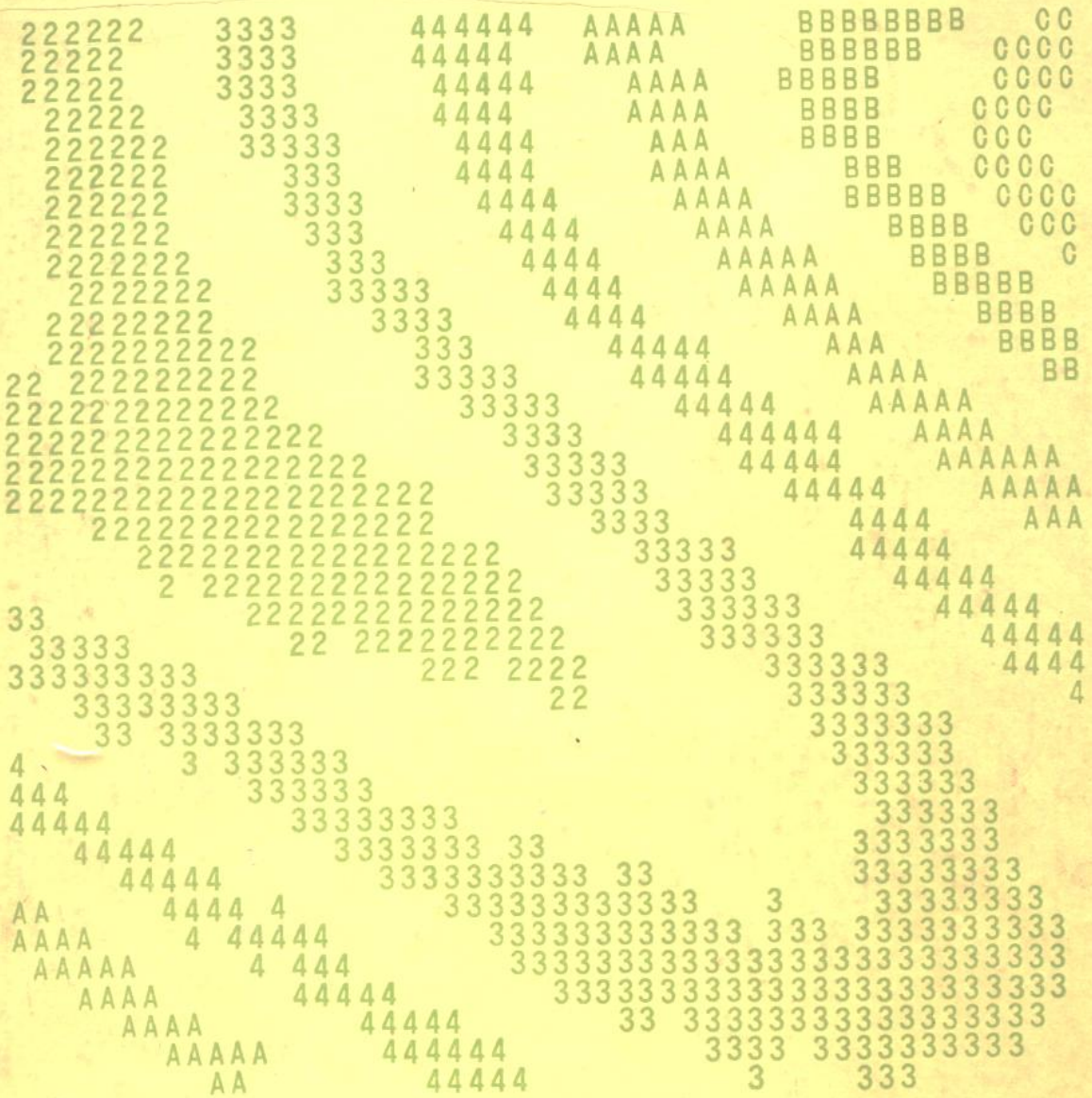
数学地质的方法与应用

地质与化探工作中的多元分析



於崇文等 编著

冶金工业出版社

数学地质的方法与应用

地质与化探工作中的多元分析

於崇文等 编著

冶金工业出版社

内 容 提 要

本书介绍数学地质的方法与应用，以地质和化探工作中所使用的各种多元统计分析方法为重点；在内容安排上，数学方法及其在计算机上的实现与实际应用并重。本书除绪论外，共分四篇，第一篇数学预备知识，第二篇多元分析方法，第三篇地质和化探工作中几种常用多元分析方法的ALGOL-60程序，第四篇多元分析在地质和化探工作中的应用，书末并有附录。第二篇除相关分析与回归分析，点群分析，判别分析，因子分析和趋势面分析等当前常用的方法之外，还包括了非线性映射，对应分析，马尔科夫概型分析和地质统计分析等新方法。第四篇通过研究较详细的实例说明在研究和解决各种地质与化探问题上多元分析方法的选择与综合运用。第三篇给出几种常用多元分析方法的计算机程序，便于读者使用。

本书可供广大野外地质与化探工作者、专业研究人员和地质院校师生学习和参考。

数学地质的方法与应用

地质与化探工作中的多元分析

於崇文等 编著

冶金工业出版社出版

（北京灯市口74号）

新华书店北京发行所发行

冶金工业出版社印刷厂印刷

787×1092 1/16 印张 59 1/2 字数1434千字

1980年4月第一版 1980年4月第一次印刷

印数00,001～5,000册

统一书号：15062·3492 定价 **7.30** 元

前　言

为了尽快地使数学地质在地质及化探工作中得到普及和应用，从大量的地质及化探资料中取得能较深刻地阐明地质现象和揭示地质-成矿过程的有用信息，从而对各种地质体在空间和时间上进行较精确的定量解释，提高找矿效果，为实现四个现代化作出贡献，我们编写了这本《数学地质的方法与应用》（着重多元分析方法）一书，供广大野外地质及化探工作者、数学地质工作者学习和参考。

全书除前言、绪论和附录外共分四篇。第一篇数学预备知识，主要为广大地质及化探工作者介绍一些应用数学地质方法所必需的数学知识。第二篇多元分析方法，对目前国内外所使用的多元分析方法逐章介绍，建立各种方法的较清晰的数学模型，使读者能较好地理解和掌握方法的实质。除对目前国内外常用的几种多元分析方法（例如相关分析及回归分析，趋势面分析，点群分析，判别分析和因子分析）作了较详尽讨论之外，还较详细地介绍了近年来在上述方法基础上发展起来的对应分析、线性映射分析、马尔科夫概型分析和地质统计分析等方法，目的是引起地质及化探工作者的注意，开阔思路，集思广益，在实际工作中逐步应用。第三篇ALGOL-60程序（DJS-21机）给出了目前我国常用的和较重要的一些多元分析方法的计算机程序，便于读者使用。第四篇是多元分析在地质和化探工作中的应用。

本书由武汉地质学院地球化学教研室和数学教研室部分同志在冶金部地质与化探多元分析学习班有关讲义的基础上进行补充、修改并重新编写而成。绪论、第二篇、第四篇和附录主要由地球化学教研室於崇文同志编写，并负责全书最后的统一整理工作；第一篇和第三篇主要由数学教研室蒋耀淞同志和王长庚同志编写；地球化学教研室沈镛立和吴悦斌同志参加了第二篇和第三篇的部分工作。

本书编写过程中编者较多地引用了国外的工作成果，也引用了国内的某些研究成果（第五章判别分析引用了中国科学院计算所杨自强同志和北京医学院方积乾同志的研究成果等），并承蒙许多同志审阅，提出了许多修改意见，有关单位在出版、绘制图稿等方面给予大力支持，编者对他们表示衷心感谢。

由于编者水平所限，书中缺点和谬误在所难免，恳请读者批评指正。

编者

一九七八年九月

目　　录

绪　　论

一、地质科学的发展趋势

从一种运动形式辩证地过渡到另一种运动形式，以及相应地从一种科学辩证地过渡到另一种科学是《自然辩证法》的中心思想之一。早在上一世纪，革命导师恩格斯就曾指出："在分子科学和原子科学的接触点上，双方都宣称与己无关，但是恰恰就在这一点上可望取得最大的成果。"几十年来地质学的发展史完全证实了恩格斯的这一预言。在生产发展的推动下，本世纪以来地质科学和基础自然科学以及先进技术相结合，在它们的结合点上产生和发展了新的边缘科学，成为地质科学发展的一个明显的趋势，其主流和基本方向是：

地质学和物理学相结合，产生了地球物理学和地球物理探矿法。

地质学和化学相结合，产生了地球化学和地球化学探矿法。

地质学和力学相结合，产生了地质力学和地质力学探矿法。

地质学和数学相结合，产生了地质数学，并且正在逐步形成地质数学探矿法。

这些新的边缘科学的产生和发展引起了地质科学的质的飞跃，使古老的地质学展示出崭新的面貌，表现出强大的生命力。

地质数学（也称"数学地质"）是运用数学理论和方法研究各种地质现象的数量关系和空间形式的科学。数学地质通过以数学模型模拟地质现象和用快速电子计算机实现复杂、大量的运算，正在引起地质科学的重大变革：

1．使地质科学从定性走向定量

数学是研究现实世界的数量关系和空间形式的一门科学。因此，数学一旦和地质科学相结合，就要求我们从具体的地质现象中抽象出理论上的数学模型；并将各种地质特征和参数数字化，用数学关系式对地质现象作精确的表达，从而使地质科学由定性的描述发展为定量的科学。

2．由确定性模型转向概率性模型

根据地质科学近数十年来的发展和当前研究工作的动向，一种带有倾向性的看法是认为在地质科学的大部分领域内，用一定形式的微分方程来描写各种地质现象，并且在满足定解条件（已知初始条件和特定的边界条件）下求出微分方程的定解、建立确定性模型（概率$P=1$）是非常困难的。大量的事实和实践经验说明，我们不能精确地预测个别地质现象的发生和演变，因而只能用概率性模型（$0<P<1$）来研究各种地质现象总体的统计规律。这就是说，地质科学的数学概括有一种从确定性模型转向概率性模型的倾向。为了适应这一方面的需要，地质数学以概率论和数理统计作为它的主要数学工具，同时运筹学、信息论、控制论和体系分析等学科也正在逐渐引起重视。

3．从一元分析向多元分析发展

为了实现国民经济对于地质科学研究成果定量化的要求，为了提高成矿预测、远景区评价和矿床储量估计的可靠性和准确度，仅仅根据单变量的观测数据已经远远满足不了要求；而必须同时对多个变量进行观测，汇集丰富的原始信息，较全面地反映所研究地质现象的各个有关侧面。五十年代电子计算机的发明和应用为实现这一目标创造了有利的条

件，促使地质数据的一元统计分析向多元分析的方向发展。

4．从观测基础上对现象的表征和描述提高到计算机上的模拟实验

地质科学定量研究的第一步是利用观测数据对地质现象进行表征和描述。第二步是对实际地质现象进行数学抽象，构成某种数学模型，选择适当的经验函数对观测数据进行拟合。第三步则是进行模拟实验，找出地质现象和地质过程的更接近实际的条件、特征和规律。

目前，多数地质数学的研究成果都以已有的少数几类经验函数为基础，所构成的数学模型有相当大的局限性。另一方面，模拟实验大多也在实验室的条件下进行，由于实验室条件的限制，模拟实验的结果往往和真实地质过程有一定的距离。

近年来发展出一种统计试验计算法〔又名蒙特卡罗（Monte-Carlo）法〕。这种方法的基本思想是人为地设计、模拟出某种合乎需要的概率论模型，使它的某些参数恰好就是所考虑的问题的解。然后利用观测数据，根据所模拟的模型，在计算机上进行计算。将计算结果和实际地质现象进行比较，如果二者的符合程度不够理想，就可以将模型进行改造，重新计算，直至达到满意的符合程度为止。由于计算机的计算速度很高，因此可以在较短的时间内进行多次的模拟实验。这种模拟实验可以同时克服经验函数和实验室条件的局限性，多快好省地得出比较理想的结果。

本书的重点是多元分析，下面转入这一方面的讨论。

二、多元分析的涵义

（一）定义

“多元分析”（Multivariate Analysis）的基本特点是研究 n 个个体的集合，每一个个体都有 p 个变量的观测值。个体的集合可以是完备的，也可以是抽自较大集合的一个样本。变量可以是连续的或不连续的，它们本身也可以是较大集合中的一个子集合。形式上，我们可以把多元分析定义为数理统计学的一个分支，它研究相依变量的集合和个体之间的相互关系。多元分析要求通过对多元数据集合的研究达到下列几个主要目的：

1．简化结构

目的是要弄清能否通过把一个相依变量的集合变换为一个新的独立变量的集合或者约简数据集合的空间维数等较简单的方法来表示所研究的复杂的多元数据集合。

2．进行分类

目的是要查明所研究的个体究竟是或多或少偶然地散布在某一变化域内，还是分别归属于不同的组别、类别或点群。

3．组合变量

分类是对个体按其相似性进行分组，同样我们也可以对变量按其相似性进行组合。

4．分析相依性（interdependence）

目的是检查变量之间的相互依赖关系，包括从独立性（或无关性）到共线性〔即一个变量是其他一些变量的线性函数（或者更一般地，是其他一些变量的非线性函数）〕之间的种种过渡关系。

5．分析依赖性（dependence）

在第4点中我们把变量按其相互关系看成全部处于同等地位，而分析依赖性的目的则是检查一个或几个变量对于其他一些变量的依赖关系，二者并非处于同等地位。

6．提出假设和检验假设。

（二）内容

根据研究对象和研究目的，我们将多元分析的内容用图解的形式表示在图1中。必须指出，从逻辑上说，应将“时间序列”列入多元分析的内容。

这里有必要讨论一下多元分析和一元分析之间的关系。今后我们将会看到，一元分析中的有些方法可以推广到多元分析的领域，例如矩统计量（平均值，方差，偏斜度等）和以它们为基础的一些方法（相关分析，回归分析，方差分析等）就是这样的例子。但是多元分析也包括一些不能由一元分析直接推广的新方法（如主成分分析，因子分析，典型相关分析，各种分类分析等）。多元分析的下列特点是造成多元分析不能完全用一元分析的推广来概括的重要原因。

1．多元分析所面临的许多实际情况并不具有通常意义下的概率性。例如，我们所研究的对象可能是个体的整个集合，或者我们所根据的数据来自已知个体的一个子集合但并非一个随机样本。因此硬要将这样的数据按经典的概率模型去处理是不恰当的。

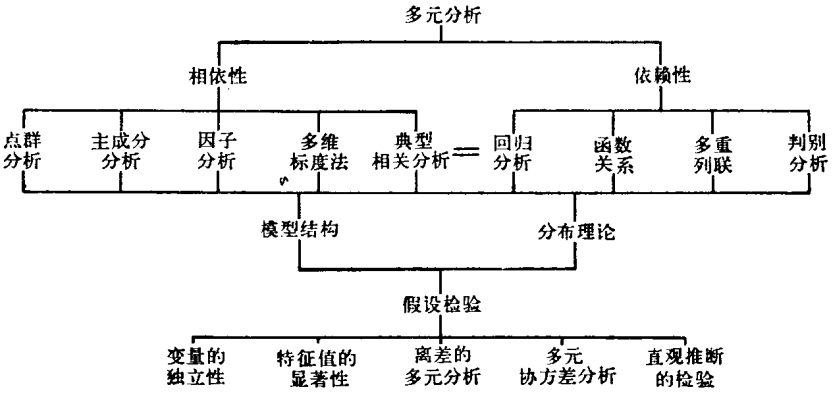


图1　多元分析内容图解

2．即使我们所研究的样本是一个随机样本，也可能因为母体分布的非正态性而使多元分析遇到困难。在单变量统计学中我们可以研究频率分布的整个变程〔像泊松分布和超几何分布等不连续型分布或者像皮尔逊曲线（Pearson curves）或埃其瓦斯展开式（Edgeworth expansions）等连续型分布〕。利用这些分布就可以详细研究非正态性对于以正态变化为基础的过程的影响。但是对于一维以上的情形，规定整个一组既可变通又附合实际的曲面就要困难得多。多元统计中的许多理论研究都是以母体服从多元正态分布的假设为基础的，而在偏离正态的条件下这些理论工作的正确性往往难以准确地估计。

3．在单变量的统计学中我们往往可以提出种种非参数性的方法，从而使这些方法具有普遍性。这些方法都依赖于有序性。在一维以上的情形，不存在线性意义下的有序性；虽然对于二维的情形还可以利用少数方法回避这一问题，然而非参数的多元方法却很少见。尤其是在2中所指出的缺乏可以用来拟合数据的整族多元分布的情形下，更增加了多元方法的困难。

4．在二维空间内用图形来表示数据是相当简单的，即使是三维空间的图示，困难也并不大。但是要想进行三维以上的图示则简直是不可能的事，散布在 p 维空间内的 n 个点

的图示只能用各种办法把它们投影到各个平面上来处理。现在已经有一些计算机程序可以将点的散布在任意一个平面上的投影显示在一个电视屏幕上，从而使我们能够从许多不同的角度来看这一散布的各个侧面。有些方法则是利用二维空间内的颜色来表示三维空间的图形。这些方法虽然对于较高维数的图示有帮助，但还不能彻底解决复杂点散布的内部结构的图示问题。

5．电子计算机的发明和应用对于多元分析的发展是一个很大的促进。现在用计算机可以在几分钟、甚至几秒钟之内完成的计算，在二、三十年前是无法完成、甚至是难以想像的。这无疑是一个巨大的进步，也是一个很大的优点；但任何事物总是一分为二的，使用计算机也包含着一些不利因素。过去，对于一些算法可以亲自进行验算，现在如果使用别人编写的程序有时就不得不信赖它。但是严格说来，我们不能盲目地接受计算机程序，而应该用一些考验的问题来检查别人编写的程序所算出的结果是否符合所需要的准确度。

（三）地质和化探工作中的多元分析方法

五十年代以来，在电子计算机应用的促进下，在多元分析本身的发展及其在其他科学领域内应用的推动下，地质和化探工作中逐步引进了多种多元分析方法，并已取得了令人鼓舞的成效。根据实际应用的现状，总的看来，可以说多元分析方法的应用已经远远超出数据处理的范畴，而正在发展成为地质和化探工作的一种富有生命力的新的研究手段，促使地质科学从定性向定量发展，引起理论上和方法上的重大变革。

近年来地质和化探工作中多元分析方法日益发展，种类甚多，我们可以根据两条根本原则去认识各种方法之间的内在联系，把握地质和化探工作中多元分析方法的发展规律。这两条原则是：

1．对立统一规律

各种地质属性的分析与综合是揭示和阐明地质现象内在联系和固有特征的基本出发点，很多多元分析方法都是研究地质属性的分析与综合的重要手段。

2．物质和运动-时间-空间三者之间的唯物辩证关系

岩体或矿床的形成既是成岩作用或成矿作用(成岩物质或成矿物质的运动)在时间上的演化过程，又是成岩作用或成矿作用在空间上的扩展和分布。地质和化探工作中应用多元分析方法的一个重要目的就是要研究地质特征的空间分布与空间相关，同时还要研究地质特征的时间变化，作为展示地质现象的变化特征和对地质现象进行成因分析的有力工具。

根据这两条根本原则大体上可以将多元分析方法分为四大类，各类中所列举的是本书将要着重介绍的几种方法。

第一类 属性的分析与综合
- 相关分析与回归分析
- 点群分析
- 判别分析
- 典型相关分析
- 因子分析
- 非线性映射分析
- 对应分析

第二类 特征的空间分布
- 趋势面分析
- 调合趋势分析
- 典型趋势面分析
- 单位向量场分析

第三类 变量的空间相关
- 地质统计分析

第四类 参数的时间变化
- 马尔科夫概型分析

（四）研究方法

虽然多元分析方法的种类很多，但无论在理论上或实践上都具有一个共同点，就是要设法简化问题的复杂性。例如，在实践上我们也许希望减少变量的数目以节省计算工作量，或者由于缺乏测试手段或费用昂贵而想要避免对某些变量进行观测，或者我们想在并不严重影响工作目的的前提下推迟对某些变量进行观测，如此等等。在理论上，我们也许愿意在损失部分信息的条件下约简空间的维数，或者我们想进行某种变换以去掉某些无意义的参数，诸如此类。

例如，假定有 p 个变量，我们想研究它们的方差和相关系数。我们需要考虑 p 个平均值，p 个方差和 $\frac{1}{2}p(p-1)$ 个相关系数，总共 $\frac{1}{2}p(p+3)$ 个参数。对于单变量，我们只需考虑两个参数；对于双变量，需要考虑五个参数；而当 $p=10$ 时，所需考虑的参数就不下65个。显然如果我们能够将变量变换为一个不相关的变量集合，则将减少 $\frac{1}{2}p(p-1)$ 个变量，从而大大简化了表示的复杂性。或者，如果我们能够将维数从 p 约简为 $(p-1)$，则即可减少 $\frac{1}{2}p(p+3)-\frac{1}{2}(p-1)(p+2)=p+1$ 个参数。

当然有时我们并不希望通过对每一个参数的估计来十分详细地表征所研究的体系。但是显然，我们能够约简的参数的数目愈多，我们也就愈加接近体系结构的更易被理解的模型。

三、观测矩阵和协方差矩阵

假定我们所考虑的变量全部都是连续的，则可将典型的数据排列成一个 $p \times n$ 矩阵：

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} \tag{1}$$

其中 x_{ij} 是第 j 个个体的第 i 个变量的值。在实践上，由于 n 往往比 p 大得多，因而在整理数据时，以矩阵的列表示变量，而以行表示个体较为方便。

计算时，将每一个变量的值对中心进行校正（或称“中心化”，即将变量的值减去它的 n 个值的平均值）往往是比较方便的。我们用一个点下标代表这个平均值，即

$$x_{i\cdot}=\frac{1}{n}\sum_{j=1}^{n}x_{ij} \tag{2}$$

如果我们将中心化变量值的矩阵右乘以它的转置矩阵，就得到一个 $p \times p$ 矩阵，其中的典型项是：

$$nc_{ij}=\sum_{k=1}^{n}(x_{ik}-x_{i\cdot})(x_{jk}-x_{j\cdot}) \tag{3}$$

其中 c_{ij} 是 x_i 和 x_j 的协方差。我们可以将此矩阵写作

$$\boldsymbol{C}=\frac{1}{n}\boldsymbol{X}\boldsymbol{X}' \tag{4}$$

其中 $\boldsymbol{X}'$ 代表 $\boldsymbol{X}$（为一 $n \times p$ 矩阵）的转置矩阵。今后，除非有专门说明，我们在定义方差

和协方差项时都用 n 而不用（$n-1$）作除数。$\boldsymbol{C}$代表协方差矩阵或离差矩阵。

特别是，如果每一个x_i除以x_i的方差的平方根，则协方差矩阵就变为相关矩阵

$$\begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{12} & 1 & r_{23} & \cdots & r_{2p} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ r_{1p} & r_{2p} & r_{3p} & \cdots & 1 \end{bmatrix} \tag{5}$$

协方差矩阵和相关矩阵都是对主对角线对称的，即$c_{ij}=c_{ji}$，$r_{ij}=r_{ji}$。

必须强调指出，协方差矩阵及其行列式在多元分析的理论中起着根本的作用，如同方差在一元分析理论中所起的作用那样。

我们对于观测矩阵（1）可以从两个角度去考虑。如果从横向（即按行）去看，比如将两行进行比较，那么我们就是在检查 n 个个体的集合中该两个变量之间的关系。如果从纵向（即按列）去看，比如将两列进行比较，那么这一比较就将给出 p 个变量的集合中该两个体之间的关系。

四、数据集合的两种空间表示

对于观测矩阵的两种对偶式的读法对应于数据的两种截然不同的几何表示方法。

（一）数据集合的第一种空间表示

这种空间表示是把数据集合看作 p 维空间中的 n 个点。它是回归分析和相关分析中的散布图的自然推广。我们想像有 p 个正交轴，确定一个 p 维空间。个体集合中的任一个体对应于 p 维空间中的一个点，第 j 点的沿第 i 轴的坐标是 x_{ij}。因而 n 个个体就对应于 p 维空间中的 n 个点，而我们注意力的集中点则是这一群点的模式：究竟它们是坍塌成一个较低维的扁平空间，还是它们或多或少呈球状分布，或者呈伸长的形状；它们是否可以清楚地分成若干组，等等。当然，为了便于表示，我们只要能画出二维的图形就可以了，但就其性质来说这个问题却具有一般性。

值得指出，这种几何的表示方法毫不依赖于三维以上空间的“存在”与否。我们可以把它看作是用一种较方便的语言来表示可以用代数术语表达的数学事实。比如，对于同一个数学事实，我们分别用如下的几何语言和代数语言表示，则显然几何语言的表达方式比较方便：

（1）一个（$p-1$）维的超平面和一个 p 维的超球面相交于一个（$p-1$）维的超球面（类似于一个平面和一个球面相交于一个元）。

（2）$\sum_{i=1}^{p}(x_i-a)^2=b$ 和 $\sum_{j=1}^{p}c_jx_j=d$

的轨迹可以经过一个适当的坐标变换表示成如下的形式

$$\sum_{i=1}^{p-1}(y_i-k)^2=l$$

（二）数据集合的第二种空间表示

这种空间表示是把数据集合看作 n 维空间中的 p 个点。

我们取 n 个正交轴，其中每一个轴对应于 n 个个体集合中的每一个个体；并确定 p 个点P_i，P_i的坐标是 n 个值：$x_{i1}-x_{i\cdot}$，…，$x_{in}-x_{i\cdot}$。因而对于每一个变量就有一个对应的点。P_i与原点O的距离由下式给出：

$$OP_i^2=\sum_{j=1}^{n}(x_{ij}-x_{i\cdot})^2 \tag{6}$$

即x_i的方差的n倍。

如果将变量按方差为1进行标准化，则p个向量OP_i都具有单位长度，并且它们的端点都落在一个单位半径的超球面上。此外，OP_i和OP_j之间所成角度θ_{ij}的余弦

$$\cos\theta_{ij}=\sum_{k=1}^{n}(x_{ik}-x_{i\cdot})(x_{jk}-x_{j\cdot})\Big/\left\{\sum_{k=1}^{n}(x_{ik}-x_{i\cdot})^2\sum_{k=1}^{n}(x_{jk}-x_{j\cdot})^2\right\}^{1/2} \tag{7}$$

是x_i和x_j之间的相关系数。

p个向量OP_1，…，OP_p确定一个嵌入在n维空间内的p维空间。我们可以把它们想象为一把雨伞的伞骨。如果两个变量高度相关，则r接近于1，并且θ_{ij}接近于零。如果两个变量为零相关，则θ_{ij}是$\frac{1}{2}\pi$，并且向量是正交的。这些向量“集拢”的程度是它们所代表的变量之间相关系数大小的度量。如果一个向量位于其它向量所确定的$(p-1)$维空间内，则它所代表的变量是其他一些变量的线性函数。这里，向量的模式揭示变量之间相互关系的性质，如同第一种空间表示中n个点的模式揭示个体之间相互关系的性质那样。

在多元分析方法的讨论中经常遇到的所谓“R式分析”（研究变量之间的相互关系）就相当于第二种空间表示，而所谓“Q式分析”（研究个体之间的相互关系）则相当于第一种空间表示。

五、空间的维数和矩阵的秩

矩阵论的基本概念之一是关于“秩”的概念。我们称一个$p\times n$矩阵的秩为m（$m\leqslant p$，n中之较小者），如果得自此矩阵的多于m行和m列的所有行列式（子式）均等于零，但至少有一个$m\times m$行列式不等于零。矩阵论的一个重要结论是：如果一个矩阵的秩是m，则所有的值x_{ij}线性依赖于它们（比如x_{ij}；$i,j=1,2,\cdots,m$）的m个集合。用几何的语言来说，在第一种空间表示中，n（$n>p$）个点将位于一个$m\leqslant p$维的空间内。

矩阵论的另一个结论是：一个矩阵与其转置矩阵的乘积的秩等于此矩阵的秩。对于我们所讨论的情形，这就等于协方差矩阵或相关矩阵的秩。因而，为了确定一个$p\times n$矩阵的秩，可以不必检查原始矩阵的所有行列式，而只要考虑$p\times p$协方差矩阵的秩。

例如，我们考虑一种四个变量的情形，设有矩阵

$$\begin{bmatrix}1 & 0.8 & 0.6 & 0.6\\ 0.8 & 1 & 0.96 & 0\\ 0.6 & 0.96 & 1 & -0.28\\ 0.6 & 0 & -0.28 & 1\end{bmatrix} \tag{8}$$

整个矩阵的行列式等于零。同样，划去一行和一列后的所有3×3行列式也都等于零。但是2×2行列式却不等于零。因而此矩阵的秩为2（$p=4$，$m=2$）。

由此可以得出结论，这一个貌似四维的变差实质上却是二维的，就是说，我们可以找到两个新变量，比如ξ_1和ξ_2，用它们来表示x_1，x_2，x_3，x_4。下面就是这样的一对新变量（不一定是唯一的）：

$$\left.\begin{aligned} x_1&=\xi_1\\ x_2&=0.8\xi_1+0.6\xi_2\\ x_3&=0.6\xi_1+0.8\xi_2\\ x_4&=0.6\xi_1-0.8\xi_2 \end{aligned}\right\} \tag{9}$$

我们不能用x_1，x_2，…唯一地表示ξ_1，ξ_2；但是反之，由（9）式，我们可以得到

$$\begin{aligned} \mathrm{var}\xi_1&=\mathrm{var}x_1=1\\ \mathrm{var}\xi_2&=\mathrm{var}\left(-\frac{4}{3}x_1+\frac{5}{3}x_2\right)\\ &=\frac{16}{9}+\frac{25}{9}-\frac{40}{9}\mathrm{cov}(x_1x_2)\\ &=\frac{41}{9}-\frac{32}{9}=1 \end{aligned}$$

因而容易证明 x 之间的相关系数事实上就是（8）式的相关系数。

现在我们考虑一个 p 维的数据集合，其中所有变量之间的相关系数都相等，相关矩阵为

$$\begin{bmatrix} 1 & r & r & \cdots & r\\ r & 1 & r & \cdots & r\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ r & r & r & \cdots & 1 \end{bmatrix} \tag{10}$$

将此矩阵的各行相加，提出因子$\{1+(p-1)r\}$，然后将单位行的 r 倍从所有其他各行中减去，则可求得此矩阵的行列式为

$$(1-r)^{p-1}\{1+(p-1)r\} \tag{11}$$

除非$r=1$或$r=1/(p-1)$，(11)式不可能等于零。除了这两种特殊情形，该矩阵的秩必须是 p。因此可以得出结论：我们不可能将一个彼此的相关系数相等的变量集合表示在一个低于 p 维的空间内。

我们再考虑对称矩阵（假定为 p 阶）的秩。这时我们不必检验每一个子式。如果有一个 m 阶的主子式（即对角线元素取自原始矩阵的对角线元素的子式）不等于零，并且（a）当任意一行和对应的一列加到该主子式时，所得的$m+1$阶子式均等于零，（b）当任意两行和对应的两列加到该主子式时，所得的$m+2$阶子式均等于零，则该对称矩阵的秩为m。

对于 p 阶行列式，我们有 $p-m$ 种可能将任意一行和对应的一列加到m阶主子式上，有$\frac{1}{2}(p-m)(p-m-1)$种可能将任意两行和对应的两列加到m阶主子式上。因而使 p 阶对称矩阵的秩为m的条件数是$\frac{1}{2}(p-m)(p-m+1)$。事实上，这些条件是独立条件。

最后简单提一下多元分析的计算中经常遇到的一个问题。不少多元分析的运算涉及到协方差矩阵的求逆，也就是它的行列式的倒数。实践中，协方差矩阵的行列式等于零的情形虽属罕见，但是取很小值的情形却相当常见。在这种情形下，特别对于维数较高的行列式，经过多次迭代就会产生累积误差，影响计算结果的精度。这时必须采取适当的措施加以处理。

六、多元分析方法的工作步骤

进行多元分析的工作步骤如下：

1．根据工作目的选择适当的多元分析方法，构成相应的数学模型。

2．确定合理的计算步骤和方法。

3．将此计算步骤和方法编成算法语言程序。

4．通过计算机实现运算。

5．对计算结果进行整理和分析。